CHAPTER 14

THE VARIATIONAL PRINCIPLE

We introduced in chapter 3 the topological entropy $h(T)$ of a continuous map $T : X \to X$ of a compact metric space X and in chapter 8 the entropy $h_\mu(T)$ of a T-invariant probability measure μ. In this chapter we show that these two notions are closely related.

14.1 The variational principle for entropy

The main result of this chapter is the following.

THEOREM 14.1 (VARIATIONAL PRINCIPLE). *Let $T : X \to X$ be a continuous map on a compact metric space.*

(1) *For any T-invariant probability measure μ we have that $h_\mu(T) \leq h(T)$.*
(2) $h(T) = \sup\{h_\mu(T)$: *μ is a T-invariant probability measure*$\}$.

14.2 The proof of the variational principle

The proof we give is due to Misiurewicz [1]. Recall that the topological entropy of a cover $\mathcal{U}$ is $H(\mathcal{U}) = \log N(\mathcal{U})$ and the entropy of a partition α with respect to μ is $H_\mu(\alpha) = -\sum_{A\in\alpha} \mu(A) \log \mu(A)$.

PROOF OF (1). Fix a finite Borel measurable partition $\alpha = \{A_1, \dots , A_k\}$ for X. Given $\epsilon > 0$, say, we want to "improve" this partition by choosing a family of closed sets $\hat{A}_1, \dots , \hat{A}_k$ such that

(1) $\hat{A}_i \subset A_i$, $i = 1, \dots , k$, and
(2) $\mu(A_i - \hat{A}_i) < \epsilon$,

and then defining a new partition $\hat{\alpha} = \{\hat{A}_1, \dots , \hat{A}_k, V\}$, where $V = X - \left(\cup_{i=1}^k \hat{A}_i\right)$.

We can consider an open cover for X defined by

$$\mathcal{U} = \left\{\hat{A}_1 \cup V, \dots , \hat{A}_k \cup V\right\}$$

If we compare the open covers $\vee_{i=0}^{n-1} T^{-i}\mathcal{U}$ and the partitions $\vee_{i=0}^{n-1} T^{-i}\hat{\alpha}$ then we see that

$$N\left(\vee_{i=0}^{n-1} T^{-i}\hat{\alpha}\right) \leq 2^n N\left(\vee_{i=0}^{n-1} T^{-i}\mathcal{U}\right), \quad n \geq 1 \tag{14.1}$$

(where we recall that $N(\vee_{i=0}^{n-1}T^{-i}\mathcal{U})$ is the number of elements in a minimal subcover for $\vee_{i=0}^{n-1}T^{-i}\mathcal{U}$ and $N(\vee_{i=0}^{n-1}T^{-i}\hat{\alpha})$ is the number of non-trivial elements in $\vee_{i=0}^{n-1}T^{-i}\hat{\alpha}$).

SUB-LEMMA 14.1.1. $H_\mu(\vee_{i=0}^{n-1}T^{-i}\hat{\alpha}) \leq \log N\left(\vee_{i=0}^{n-1}T^{-i}\hat{\alpha}\right)$.

PROOF. Assume that $\vee_{i=0}^{n-1}T^{-i}\hat{\alpha} = \{C_1, \dots, C_N\}$; then we can write $H_\mu(\vee_{i=0}^{n-1}T^{-i}\hat{\alpha}) = -\sum_{i=1}^{N}\mu(C_i)\log\mu(C_i)$.

■

We can use Sub-lemma 14.7 to bound

$$\begin{aligned} &H_\mu(\vee_{i=0}^{n-1}T^{-i}\hat{\alpha}) \\ &\leq \log N\left(\vee_{i=0}^{n-1}T^{-i}\hat{\alpha}\right) \\ &\leq n\log 2 + \log N\left(\vee_{i=0}^{n-1}T^{-i}\mathcal{U}\right) \qquad \text{(by (14.1))}. \end{aligned}$$

Recalling that

$$h(T) \geq h(T,\mathcal{U}) = \lim_{n\to+\infty}\frac{1}{n}H(\vee_{i=0}^{n-1}T^{-i}\mathcal{U})$$

and

$$h_\mu(T,\alpha) = \lim_{n\to+\infty}\frac{1}{n}H_\mu(\vee_{i=0}^{n-1}T^{-i}\alpha)$$

we see that $h_\mu(T,\hat{\alpha}) \leq \log 2 + h(T)$. Moreover, by Corollary 8.6.1 we have that

$$\begin{aligned} |h_\mu(T,\hat{\alpha}) - h_\mu(T,\alpha)| &\leq H_\mu(\alpha|\hat{\alpha}) + H_\mu(\hat{\alpha}|\alpha) \\ = -\sum_{C\in\alpha}\sum_{\hat{C}\in\hat{\alpha}}&\mu(C\cap\hat{C})\log\left(\frac{\mu(C\cap\hat{C})}{\mu(\hat{C})}\right) \\ -\sum_{C\in\alpha}\sum_{\hat{C}\in\hat{\alpha}}&\mu(C\cap\hat{C})\log\left(\frac{\mu(C\cap\hat{C})}{\mu(C)}\right) < 1, \end{aligned}$$

say, providing ϵ was sufficiently small.

Since α was arbitrary, we see that

$$h_\mu(T) = \sup\{h_\mu(T,\alpha) : \alpha \text{ is a finite partition}\} \leq h(T) + \log 2 + 1.$$

Finally, we can apply the argument to iterates T^k $(k \geq 1)$ to see that $h_\mu(T^k) \leq h(T^k) + \log 2 + 1$. By Corollary 3.8.1 we know that $h(T^k) = kh(T)$. The following gives the analogous result for measure theoretic entropy.

SUB-LEMMA 14.1.2 (ABRAMOV'S THEOREM). *For* $k \geq 1$, $h_\mu(T^k) = kh_\mu(T)$.

PROOF. Given any partition α we observe that

$$\begin{aligned} h_\mu\left(T^k, \vee_{i=0}^{k-1}T^{-i}\alpha\right) &= \lim_{n\to+\infty}\frac{1}{n}H_\mu\left(\vee_{i=0}^{n-1}T^{-ik}\left(\vee_{j=0}^{k-1}T^{-j}\alpha\right)\right) \\ &= \lim_{N\to+\infty}\frac{k}{N}H_\mu\left(\vee_{i=0}^{N-1}T^{-i}\alpha\right) = kh_\mu(T,\alpha). \end{aligned}$$

Given $\epsilon > 0$ we can choose α with $h(T,\alpha) > h_\mu(T) - \epsilon$ so that we have

$$\begin{aligned} h_\mu(T^k) &\geq h_\mu\left(T^k, \vee_{i=0}^{k-1} T^{-i}\alpha\right) \\ &\geq k h_\mu(T,\alpha) \geq k h_\mu(T) - k\epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary we see that $h_\mu(T^k) \geq k h_\mu(T)$.

To get the reverse inequality, notice that $h_\mu\left(T^k,\alpha\right) \leq h_\mu\left(T^k, \vee_{i=0}^{k-1} T^{-i}\alpha\right)$, using Lemma 8.6. Given $\epsilon > 0$ we can choose α with $h_\mu(T^k,\alpha) > h_\mu(T^k) - \epsilon$ and then

$$\begin{aligned} k h_\mu(T) \geq k h_\mu(T,\alpha) &= h_\mu\left(T^k, \vee_{i=0}^{k-1} T^{-i}\alpha\right) \\ &\geq h_\mu\left(T^k,\alpha\right) > h_\mu(T^k) - \epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary we see that $h_\mu(T^k) \leq k h_\mu(T)$.

■

We can now complete the proof of (1) since

$$\begin{aligned} h_\mu(T) &= \lim_{k\to+\infty} \frac{h_\mu(T^k)}{k} \\ &\leq \lim_{k\to+\infty} \frac{h(T^k)}{k} + \lim_{k\to+\infty} \frac{\log 2 + 1}{k} = h(T). \end{aligned}$$

■

Proof of (2). It suffices to show that given $\delta > 0$ there exists a T-invariant probability measure μ with $h_\mu(T) \geq h(T) - \delta$. We want to choose $\epsilon > 0$ sufficiently small that $\lim_{n\to+\infty} \frac{1}{n}\log(s(n,\epsilon)) \geq h(T) - \delta$, where $s(n,\epsilon)$ is the maximal cardinality of an (n,ϵ)-separating set. We can find a sub-sequence $n_i \to +\infty$ such that $\frac{1}{n_i}\log(s(n_i,\epsilon)) = h(T)$. Let S_{n_i} be such an (n_i,ϵ)-separated set.

For each n_i we can define a (possibly non-invariant) probability measure

$$\nu_{n_i} = \frac{1}{s(n_i,\epsilon)} \sum_{x\in S_{n_i}} \delta_x.$$

In order to arrive at a T-invariant probability measure we can consider an accumulation point μ (in the weak-star topology) of the measures

$$\mu_{n_i} = \frac{1}{n_i} \sum_{r=0}^{n_i-1} (T^r)^* \nu_{n_i}.$$

By replacing $\{n_i\}$ by a sub-sequence, if necessary, we can assume that $\mu_{n_i} \to \mu$.

Let want to consider a finite partition $\alpha = \{A_1, \ldots, A_k\}$ such that

(1) $\operatorname{diam}(A_i) < \epsilon$, $i = 1, \ldots, k$; and
(2) $\mu(\partial A_i) = 0$, for $i = 1, \ldots, k$.

Since S_{n_i} is an (n_i,ϵ)-separated set we know that each set $C \in \alpha^{(n_i)} := \vee_{j=0}^{n_i-1} T^{-j}\alpha$ contains at most one point $x = x_C \in S_{n_i}$. Thus of the sets in S_{n_i} there are $s(n_i,\epsilon)$ sets with ν_{n_i}-measure $\frac{1}{s(n_i,\epsilon)}$ and the remainder have ν_{n_i}-measure zero. In particular, we see that

$$\log(s(n_i,\epsilon)) = -\sum_{C\in\alpha^{(n_i)}} \nu_{n_i}(C)\log\nu_{n_i}(C). \tag{14.2}$$

In order to take limits in a sensible way we fix first $1 < N < n_i$ and then $0 \le j \le N-1$. We can write

$$\alpha^{(n_i)} = \vee_{i=0}^{n_i-1} T^{-i}\alpha = \left(\vee_{\substack{l=j \pmod{N} \\ 0\le l\le n_i-N}} T^{-l}\left(\vee_{i=0}^{N-1} T^{-i}\alpha\right)\right) \vee \left(\vee_{i\in E} T^{-i}\alpha\right)$$

where $E = \{0,1,\dots,j-1\} \cup \{M_j, M_j+1,\dots,n_i-1\}$, with $M_j = N\left[\frac{n_i-j}{N}\right]$, has cardinality at most $2N$.

SUB-LEMMA 14.1.3. *Given measurable partitions β and γ we have that*

$$H_{\nu_{n_i}}(\beta\vee\gamma) \le H_{\nu_{n_i}}(\beta) + H_{\nu_{n_i}}(\gamma)$$

PROOF. For *invariant* measures, this would be an immediate consequence of Lemma 8.4 (and Corollary 8.4.1). However, although in chapter 8 we assumed that the ambient measures were invariant, this property was not used at this stage and the result remains true without it.

■

In particular, we have that

$$\begin{aligned}
&-\sum_{C\in\alpha^{(n_i)}} \nu_{n_i}(C)\log\nu_{n_i}(C) \\
&\le \sum_{\substack{l=j \pmod{N} \\ 0\le l\le N-n_i}} \left(-\sum_{C\in T^{-l}\alpha^{(N)}} \nu_{n_i}(C)\log\nu_{n_i}(C)\right) \\
&\qquad + \sum_{i\in E}\left(-\sum_{C\in T^{-i}\alpha^{(N)}} \nu_{n_i}(C)\log\nu_{n_i}(C)\right) \\
&\le \sum_{r=0}^{M_j}\left(-\sum_{D\in\alpha^{(N)}} (T^{rN+j})^*\nu_{n_i}(D)\log((T^{rN+j})^*\nu_{n_i}(D))\right) \\
&\qquad + 2N\log k
\end{aligned} \tag{14.3}$$

(where for $l = rN + j$ there is a natural correspondence between $D \in \alpha^{(N)}$ and $C \in T^{-l}\alpha^{(N)}$ with $(T^l)^*\nu_{n_i}(D) := \nu_{n_i}(T^{-l}D) = \nu_{n_i}(C)$ and $C = T^{-l}D$). Summing the inequalities (14.3) over $j = 0, \dots, N-1$ we have by (14.2)

$$
\begin{aligned}
&N \log(s(n_i, \epsilon)) \\
&\leq \sum_{l=0}^{n_i-1} \left(- \sum_{D \in \alpha^{(N)}} (T^l)^*\nu_{n_i}(D) \log((T^l)^*\nu_{n_i}(D)) \right) + 2N^2 \log k
\end{aligned} \tag{14.4}
$$

Sub-lemma 14.1.4. *Let α be a measurable partition and let ν_1 and ν_2 be (not necessarily invariant) probabilty measures; then given $0 \leq a \leq 1$ we have that*

$$
\begin{aligned}
&\sum_{A \in \alpha} [a\nu_1 + (1-a)\nu_2](A) \log[a\nu_1 + (1-a)\nu_2](A) \\
&\leq a \left(\sum_{A \in \alpha} \nu_1(A) \log \nu_1(A) \right) + (1-a) \left(\sum_{A \in \alpha} \nu_2(A) \log \nu_2(A) \right).
\end{aligned}
$$

Proof. This follows immediately since $t \mapsto t \log t$ is convex.

■

Dividing (14.4) by $n_i N$ we get that

$$
\begin{aligned}
&\frac{1}{n_i} \log(s(n_i, \epsilon)) \\
&\leq \frac{1}{n_i} \sum_{r=0}^{n_i-1} \left(-\frac{1}{N} \sum_{D \in \alpha^{(N)}} (T^r)^*\nu_{n_i}(C) \log((T^r)^*\nu_{n_i}(C)) \right) + \frac{2N \log k}{n_i} \\
&\leq -\frac{1}{N} \sum_{C \in \alpha^{(N)}} \mu_{n_i}(C) \log \mu_{n_i}(C) + \frac{2N \log k}{n_i}
\end{aligned}
$$

where we have used Sub-lemma 14.1.4 repeatedly for the last line.

Since we have assumed $\mu(\partial A_i) = 0$, letting $n_i \to +\infty$ (with N fixed) we have that

$$
- \sum_{C \in \alpha^{(N)}} \mu_{n_i}(C) \log \mu_{n_i}(C) \to H_\mu(\alpha^{(N)}).
$$

This means that

$$
\begin{aligned}
h(T) - \delta &\leq \lim_{n_i \to +\infty} \frac{1}{n_i} \log(s(n_i, \epsilon)) \\
&\leq \frac{1}{N} H_\mu(\alpha^{(N)}) + \lim_{n_i \to +\infty} \frac{2N^2 \log k}{n_i} \\
&= \frac{1}{N} H_\mu(\alpha^{(N)}).
\end{aligned}
$$

Letting $N \to +\infty$ we have that

$$h(T) - \delta \leq \lim_{N \to +\infty} \frac{1}{N} H_\mu(\alpha^{(N)}) = h_\mu(\alpha) \leq h_\mu(T).$$

Since $\delta > 0$ is arbitrary this completes the proof.

■

14.3 Comments and reference

The proof we give is due to Misiurewicz [1]. Theorem 14.1 (1) was originally due to Goodman. Theorem 14.1 (2) was subsequently proved by Walters.

References

1. M. Misiurewicz, *A short proof of the variational principle for a* $\mathbb{Z}_+^N$ *action on a compact space*, Astérisque **40** (1976), 147-187.